

Improving the Automated Discovery of Clinical Trials via Web Services: Progress on a Standards Initiative of the Association of American Cancer Institutes

Lawrence B. Afrin, M.D.^a; Richard M. Evans, Jr., B.A.^b; Kevin Mitchell, M.S.^c;
Richard M. DeTeresa, B.S.^d; Dave Ko, B.S.^e; and Douglas C. Stahl, Ph.D.^e

^aHollings Cancer Center, Medical University of South Carolina, Charleston, SC

^bUSC/Norris Comprehensive Cancer Center, Los Angeles, CA

^cUniversity of Pittsburgh Cancer Institute, Pittsburgh, PA

^dMoore's UCSD Cancer Center, La Jolla, CA

^eCity of Hope Comprehensive Cancer Center, Duarte, CA

Abstract

Discovering available clinical trials is challenging and a recognized barrier to patient enrollment. In 2002 the Association of American Cancer Institutes' (AACI) Technology Task Force developed a rudimentary web service specification for programmatic discovery of trials. At the time of our initial report in 2003, five centers had deployed compliant services and demonstrated their ability to exchange basic clinical trial information. Since that time, the task force has advanced this work on a number of fronts including:

- Recruitment of additional cancer centers which have deployed compliant services;
- Extension of the AACI-compliant service model to support interaction with non-compliant trial information repositories;
- Expansion of the service specification to support a broader range of trial data elements;
- Automated discovery of AACI-compliant services via Universal Data Discovery and Integration (UDDI); and
- Development of new and improved service-based applications.

This paper details these advances and describes future plans.

Background

As described in our initial report¹, most clinical trial centers today manage information electronically, some developing their own databases, some using commercial solutions. Database schemas for such systems range in complexity, but most contain a

common set of basic elements describing each available trial (e.g., protocol title, accrual status, *etc.*). These databases are used primarily for local, center-specific trial management, but there is also substantial demand for this information from patients, investigators, sponsors, and other stakeholders.

Meeting this demand for accurate and timely trial information is a challenging task for trial centers. Likewise, finding trial information is a challenging task for information consumers.² Centers have undertaken a wide variety of methods to publicize their trials, but the continuing difficulties associated with discovery of basic trial information speak to the inadequacies of these methods. An optimal trial search method would examine all trial databases in parallel, maximizing the currency and the scope of the search while avoiding central points of failure. However, implementation of such a method requires global adoption of at least two sets of standards: (1) a standard format for *reporting* trial information, and (2) standard methods for programmatically *finding* trial databases and *communicating* queries and reports about trials. The Technology Task Force of the Association of American Cancer Institutes (AACI) has developed candidate standards as well as tools demonstrating their utility.

The history of this initiative is detailed in our previous paper.¹ In summary, a collaboration within the AACI produced an initial set of web service-based specifications for publishing and discovering basic data elements in an arbitrary trial database. A client can use this approach to discover trials listed in any trial database that complies with the service specifications. Compliant services return XML documents with consistent structure, permitting each client to obtain and consolidate clinical trial

information from disparate databases. Initially, five AACI centers developed such web services, with each center using its own preferred development environment (e.g. Visual Basic .NET, Java, ColdFusion). A prototype application that serially consumed these services and produced a dynamically generated, consolidated list of trials from all five centers was made publicly available.

New Developments

Since its original report, the task force has advanced its work on a number of fronts as detailed below.

Recruitment of Additional Centers

Since our original report, two additional AACI member centers (Oregon Health & Science University Cancer Institute and the University of Colorado Cancer Center) have deployed AACI-compliant Clinical Trial Discovery Web Services. A table listing the seven centers currently offering compliant services, and the addresses of the machine and user interfaces of those services, is available on the Internet.³ These centers, including some National Cancer Institute-designated centers, constitute

approximately 10% of the current AACI membership and are reflective of the general diversity of the membership in many aspects such as location (see Figure 1), age, and scope and volume of trial activity. Service code developed by the participating centers for the Microsoft .NET, Macromedia ColdFusion, and Java environments is also available on-line.⁴

Extension of the AACI-Compliant Service Model to Support Interaction with Non-Compliant Trial Information Repositories

Centers that have developed their own web service interfaces have had the advantage of ready access to the internals of their trial management systems and the ability to interact directly with their local databases. In contrast, center using commercial trial management systems often have only limited access to system internals. The “limited internal access problem” also applies to users of commercial (e.g., CenterWatch) and non-commercial (e.g., ClinicalTrials.gov) clinical trial registries. Some registries, such as those featured on websites for patient advocacy organizations, contain relatively few trials, while others contain thousands. Most of these registries do not yet offer programmatic interfaces for machine-based searching. The few



Figure 1. Trial Centers and Registries Accessible via the AACI Clinical Trial Discovery Web Service.

registries that do offer such interfaces have not, to date, provided trial information via a standardized format and method because such standards did not heretofore exist. It would be regrettable if the trials information contained in commercial systems and trial registries were excluded from the universe of trials discoverable via the AACI standard.

Fortunately, access to trial information contained in such systems can be achieved by developing an AACI-compliant service that interacts with one or more of the non-compliant system's existing methods for exposing trial information. In some cases this may require primitive "screen-scraping," whereas in other cases it may be accomplished via use of vendor-provided application programming interfaces (APIs).

To model such an approach for interacting with commercial trial management systems, we developed an AACI-compliant service that communicates with the non-AACI-compliant web service available in a commercial system (Oncore, PercipEnz Technologies, Inc.) currently used by ten AACI member centers (Figure 1). This interface more than doubled the number of centers with trials discoverable via AACI-compliant web services.

To model this approach for clinical trial registries, we developed an AACI-compliant web service that leverages the XML-based programmatic interface to the ClinicalTrials.gov database. As a result, the number of trials discoverable via AACI-compliant web services immediately increased by several thousand and included trials for diseases other than cancer.

Expansion of the Service Specification to Support a Broader Range of Clinical Trial Data Elements

The original version (1.0) of the AACI Clinical Trial Discovery web service provided a single method (getProtocols). Input was a search keyword string; output was an XML document containing a few key data elements (local trial identifier, protocol title, etc.) for each matching trial.

Most centers store far more data elements about their trials than were exposed by our initial service. After reviewing several available vocabularies of trial-related data elements and drawing upon the experience of task force members, we developed a much more robust description of a clinical trial, encoded as an XML Schema Document (XSD) to define the next iteration (version 2.0) of our service

specification. The XSD is available for review online.⁴

Whereas the version 1.0 specification provided only eight data elements about each trial, version 2.0 defines more than 120 data elements including trial identifier codes, trial titles, sponsors, summaries, statuses, phase, design, intent, eligibility criteria, interventions, and sites. The new specification also handles service version control issues.

As of this writing, four of the five original participating centers have deployed version 2.0-compliant services. Services providing access to the Oncore and ClinicalTrials.gov databases are also compliant with version 2.0.

Automation of AACI-Compliant Service Discovery

The demonstration applications originally developed to consume version 1.0 services had to be coded with explicit information about the locations of these services. This approach for locating target services is not scalable.

Widespread use of the Internet today is feasible in part because a client can use Domain Name System (DNS) servers to find host systems by name. Similarly, in the realm of web services, a client can use Universal Data Discovery & Integration (UDDI) servers to find locations of a named service. UDDI, a global standard developed by the Organization for the Advancement of Structured Information Standards (OASIS), enables the programmatic registration and discovery of web services. UDDI registries, or "operators," can expose basic UDDI functionality via simple access mechanisms such as URL-parameterized HTTP GET transactions. However, the full range of functionality implemented by UDDI operators is accessed via web service operations defined in the UDDI protocol. UDDI functions include adding information about a web service to a UDDI registry and updating, deleting, or retrieving such information. The UDDI protocol also specifies a replication mechanism so that service information registered with one operator is automatically replicated to other operators. In this fashion, UDDI theoretically permits discovery, through any operator, of information about any registered service, just as DNS permits discovery, through any DNS server, of the IP address of any registered domain name.

At the time of this writing, six AACI-compliant services – including the services interfaced to Oncore and ClinicalTrials.gov – have been registered with UDDI. A number of demonstration tools for

communicating with AACI-compliant services via UDDI servers have been developed and are freely available (see below).

Development of New and Improved Applications

The original demonstration search tool queried web services sequentially. Although this web-based “consolidator” tool served as proof of the concept that a common set of data elements could be drawn from multiple disparate trial databases in real-time, its serial approach limited the tool’s utility. Even with only five participating centers, the time required to query all five databases often exceeded one minute. The lack of any visual indication of query progress made the situation even more undesirable. User intolerance for long computing operations is well established.⁵⁻⁶

Since our original report, the consolidator has been enhanced to understand versions 1.0 and 2.0 of the web service specification, to query all services in parallel, and to provide a running elapsed time display for each service being queried. The source code for this Java-based application is available on the project website.⁷

With the shift from serial to parallel queries, search completion time has diminished exponentially and now seems within a range, and with accompanying visual aesthetics, likely acceptable to most users. The source code availability should help others seeking to develop efficient trial searching services.

In related work, a new, .NET-based desktop client was developed which provides consolidator functionality but also uses UDDI to discover available AACI-compliant services. This client also permits users to register new services with UDDI operators. Source code is available at the project site.⁴ A third tool demonstrating automated discovery of AACI-compliant services via UDDI was developed in ColdFusion and is also available on-line.⁸

The applications described in this section are intended to be demonstration tools and utilities, of interest primarily to informatics specialists. Our work would be nothing more than an academic curiosity without applications that provide practical benefits to end-users. Toward this end, some participating centers are developing enhancements to their local clinical trial websites that allow visitors to execute searches across all trials discoverable via AACI-compliant web services. In essence, these sites are placing consolidator applications into production use.

Significance and Future Directions

We believe our work is significant for two principal reasons. First, we have established a standard method by which trial centers can programmatically publish the information they wish to expose about their trials for discovery and use by an arbitrary information consumer. Second, we have established a draft standard format to which this information should adhere when transported from machine to machine.

The method and format are largely independent of one another. Our implementations of the method establish the “plumbing” by which arbitrary content can be programmatically communicated. Version 1.0 of our XSD provided information consumers a small amount of content in a simple format and served primarily to verify the correct operation of the plumbing. Since our original report, enhancements of the method and the format have proceeded independently. The method has been enhanced with UDDI and version control functionality, and the format has been enhanced to describe many additional clinical trial data elements. We believe that the value of the specification will increase as its scope and granularity increase, and we are working to improve these aspects. For example, information about statistical aspects of trials may be helpful to include, as would more granular information about eligibility criteria and interventions.

We are aware that relevant vocabulary standards work is currently being pursued by a number of investigators and organizations, including Health Level Seven, the Clinical Data Interchange Standards Consortium, and the NCI. The NCI has created a cancer data standards repository (caDSR) compliant with the ISO 11179 metadata repository standard. The caDSR defines common data elements (CDEs) containing metadata relevant to cancer clinical trials.

The evolving enhancement of any given area of our specification is likely to intersect with the standards work being conducted by many other organizations, and the flexibility of our approach should permit continued updates of our format independent of our method. Similarly, the increasing richness of our format may stimulate ideas for new methods of using the available data, and these methods can be developed independent of changes in the format.

A continuing primary emphasis will be to increase participation in the current project. To date, seven centers have deployed a total of nine AACI-

compliant services. The resulting “grid” of trial databases contains more than 10,000 trials being conducted across hundreds of centers. However, there likely are thousands of other trials being conducted at hundreds of other centers which are not yet programmatically accessible except perhaps via specialized techniques for probing what has been called the “deep,” “invisible,” or “hidden” web.⁹ Our work is not specific to the domain of cancer, and there is much to be gained if trial centers and organizations in other domains deployed our service. All implementations developed to date are freely available for rapid adaptation by other centers.

Another area for continued development involves methods for uniquely identifying each clinical trial. Many trials are conducted at multiple sites, each of which assigns a site-specific primary identifier. When centers do include the sponsor’s identifier in their local databases, they may be named and formatted inconsistently. This situation leads to a significant problem when using our service to discover trials at multiple sites, since different primary and secondary identifiers may reference the same trial. This situation can cause an individual multi-center trial to appear as many different trials; such “noise” may decrease the service’s utility to some consumers. Furthermore, this problem will persist regardless of whether any of the present candidates for a global trial identifier system ever achieve universal acceptance. Achieving a high signal-to-noise ratio requires the ability to identify how likely it is that two trials with different identifiers are the same trial simply being performed at two sites. This pattern-matching problem has been addressed in other domains (e.g., patient registration), and we plan to develop a similar approach in which the uniqueness of each trial is based on an evaluation of identifiers, titles, sponsors, medical domains, interventions, and other elements. The solution of this problem will be a critical step toward a higher quality clinical trial search system.

Production of additional end-user applications is also an important area for future development. For example, the service specification can be enhanced to support additional coding for eligibility criteria and interventions. A real-time Really Simple Syndication (RSS) feed of updated trial information also could be developed.

In conclusion, we have developed an improved specification, which we propose as a draft standard for the clinical research community, for a clinical trial discovery web service. Seven cancer centers, operating in a variety of computing environments,

have implemented the service against nine clinical trial information repositories containing over 10,000 trials in cancer and non-cancer domains. Further work will include (1) increasing the number of publicly discoverable, standards-based trial data elements, (2) improving search quality through unique trial identification, (3) facilitating more widespread participation in the continuing development of this global resource for all those affected by trials, and (4) development of more end-user applications.

Acknowledgements

The authors would like to express their gratitude to Srinu Kalluri and Dominic Komareddy of PercipEnz Technologies, Inc. for their efforts on behalf of their AACI member clients to make the Oncore system accessible via the web service described herein.

References

1. Stahl DC, Evans RM, Afrin LB, DeTeresa RM, Ko D, Mitchell K. Web Services-Based Access to Local Clinical Trials Databases: A Standards Initiative of the Association of American Cancer Institutes. *Proc AMIA* 2003.
2. Cassileth B. Clinical Trials: Time for action. *J Clin Onc* 2003 Mar 1; 21(5); 765-766.
3. Table of AACI Technology Task Force Team Members, http://opidev.upmc.edu/aaci/team_members/TeamMembers.jsp.
4. AACI Web Service Source Code Repository, <http://www.infosci.coh.org/aacitech/>.
5. Shneiderman B. Chapter 10: Response Time and Display Rate. In Shneiderman B, Plaisant C, eds., *Designing the User Interface*, Third Edition. Addison Wesley, ISBN 0321200586.
6. Barber RE, Lucas HC. System response time, operator productivity, and job satisfaction. *Communications of the ACM* 1983 Nov; 26(11):972-86.
7. AACI Project Website, <http://opidev.upmc.edu/aaci/project.jsp>.
8. Hollings Cancer Center Clinical Trial Discovery Web Service Project Site, http://hcc.musc.edu/research/clinical/uddi_test.cfm.
9. Chang KC, He B, Li C, *et al.* Structured Databases on the Web: Observations and Implications. Univ. of Illinois at Urbana-Champaign Technical Report, 2003 Feb, <http://eagle.cs.uiuc.edu/tr/UIUCDCS-R-2003-2321.pdf>.